

THE LAW OF SERIES

T. DOWNAROWICZ¹ AND Y. LACROIX

May 16, 2008

ABSTRACT. We consider an ergodic process on finitely many states, with positive entropy. Our first main result asserts that the distribution function of the normalized waiting time for the first visit to a small (i.e., over a long block) cylinder set B is, for majority of such cylinders and up to epsilon, dominated by the exponential distribution function $1 - e^{-t}$. That is, the occurrences of so understood “rare event” B along the time axis can appear either with gap sizes of nearly exponential distribution (like in the independent Bernoulli process), or they “attract” each-other. Our second main result states that a *typical* ergodic process of positive entropy has the following property: the distribution functions of the normalized hitting times for the majority of cylinders B of lengths n' converge to zero along a sequence n' whose upper density is 1. The occurrences of such a cylinder B “strongly attract”, i.e., they appear in “series” of many frequent repetitions separated by huge gaps of nearly complete absence.

These results, when properly and carefully interpreted, shed some new light, in purely statistical terms, independently from physics, on a century old (and so far rather avoided by serious science) common-sense phenomenon known as *the law of series*, asserting that rare events in reality, once occurred, have a mysterious tendency for untimely repetitions.

INTRODUCTION

We study the distribution functions of the hitting (and automatically also return) time statistics for small cylinder sets in processes on finitely symbols. We refer the reader to the rich literature on the subject (e.g. [A-G], [C], [C-K], [D-M], [H-L-V], [L] and the reference therein) for the recent developments in this field. Many works concentrate on determining whether a process (or a class of processes) has “exponential asymptotics” or not. These attempts were successful in rather restricted classes of processes. Our Theorem 1 (and its variant, Theorem 3) is the first fully general result saying something concrete about all ergodic positive entropy processes, from this point of view. Namely, we prove that in such processes any essential limit distribution function for the hitting times is majorized by the exponential law $1 - e^{-t}$. In particular, this excludes many behaviors proved to exist in zero entropy, such as the presence of an essential limit law for the return times concentrated away from zero.

1991 *Mathematics Subject Classification.* 37A50, 37A35, 37A05, 60G10.

Key words and phrases. stationary random process, positive entropy, return time statistic, hitting time statistic, repelling, attracting, limit law, the law of series, typical property.

This paper was written during the first author’s visits at CPT/ISITV in 2005 and 2006, supported by CNRS and ISITV. Research of the first author is supported from resources for science in years 2005-2008 as research project (grant MENII 1 P03A 021 29, Poland)

¹corresponding author: downar@im.pwr.wroc.pl

This theorem sheds a new light on the extensively studied class of ergodic processes with positive entropy, where one could expect, all general properties have been established already long ago. It is impossible not to mention here the theorem of Ornstein and Weiss [O-W2] which relates the return times of long blocks to entropy. However, this theorem says nothing about the asymptotics of the distribution of the return times, because the logarithmic limit appearing in the statement is insensitive to the proportions between the gap sizes.

Our approach is slightly different from the one represented in most papers on the return/hitting time asymptotics, as we are not interested in computing the limit laws “at points”, i.e., along cylinders shrinking to a point x , where x usually belongs to a positive (or full) measure set. We describe the restrictions on the distributions valid for “majority” of long cylinders B . The passage from our approach to the limit laws at points is described in the last section.

The proof of Theorem 1 is rather complicated, yet entirely contained within the classics of ergodic theory; it relies on basic facts on entropy for partitions and sigma-fields, some elements of the Ornstein theory (ϵ -independence), the Shannon-McMillan-Breiman Theorem, the Ornstein-Weiss Theorem on return times, the Ergodic Theorem, basics of probability and calculus.

Our Theorem 2 belongs to the category describing typical (or generic) properties. It states that a typical ergodic process with positive entropy (see the last paragraph of this section for the meaning of typicality among positive entropy processes) has the following property which we call *strong attracting*: there exists a subsequence of lengths (n') of upper density 1 in \mathbb{N} , such that the distribution functions of the normalized hitting times for the majority of cylinders B of lengths n' are “flat”, i.e., close to zero on a long interval. Recall that only not long ago ([C-K]) it was discovered that some mixing (but still of entropy zero) transformations admit nonexponential asymptotics. Our result shows that even some Bernoulli processes do so, which, in particular, answers in the negative a question of Zaqueu Coelho [C].

Both inequalities between the distribution function of the normalized hitting time for an event B and the exponential law $1 - e^{-t}$ have nice and clear interpretations in terms of what we call *attracting* – the tendency of the occurrences of B to appear in series, and *repelling* – the opposite tendency, toward a more uniform distribution of occurrences along the time axis. To our knowledge, these interpretations have not been addressed or discussed in any papers in the field. In these terms, our results can be expressed as follows: Theorem 1 – in any positive entropy process the repelling of almost every sufficiently long cylinder B is at most marginal; Theorem 2 – within any measure-preserving system of positive entropy, if we “draw” a finite partition, then most likely it will generate a process, where nearly all long blocks of certain lengths (belonging to a large subset of \mathbb{N}) strongly attract.

If we extrapolate this to processes and rare events running in reality, we obtain an astonishing contribution to the century old discussion about the so-called *law of series* (see the next section for more details).

Our understanding of typicality is somewhat different from the often considered setup, in which the set of all measure-preserving transformations (the automorphism group) on a fixed probability space is endowed with the topology of the weak convergence. In this setup, a typical transformation has entropy zero ([Ro]). Besides, the property we want to examine (strong attracting) depends on the generating partition, so we need to allow the partition to vary. Thus, we fix a measure-preserving system of positive entropy and $m \geq 2$, we consider all factor-processes

generated by varying partitions into at most m elements, and we adopt the notion of typicality with respect to the usual Rokhlin metric for partitions (which is complete on such partitions). Here, a typical process has positive entropy, so this approach is reasonable for studying “typical properties of positive entropy systems”. Although we define typicality within a fixed system, strong attracting turns out to be typical inside every positive entropy system, which makes our notion of typicality for this property universal.

Acknowledgments

The authors would like to thank Dan Rudolph for a hint leading to the first example of a positive entropy process with attracting, and, in effect, to the discovery of the attracting/repelling asymmetry. (The same basic idea is still used in the proof of Theorem 2.) We are grateful to Jean-Paul Thouvenot for his interest in the subject, substantial help, and the challenge to find a purely combinatorial proof of Theorem 1 (which we save for the future).

THE COMMON SENSE *LAW OF SERIES* VERSUS OUR RESULTS

A “series” is noted in the every-day life, when a random event considered extremely rare happens more than once in a relatively short period of time. In the common sense, the *law of series* asserts that such series occur more often than they intuitively should, indicating the existence of an unexplained physical force or statistical rule provoking them. For example, runs of good luck happen to gamblers, leading to high winnings (see [Wi] for the famous case of Charles Wells), people experience repetitions of similar unlucky events (hence the proverb “misfortune never comes alone”), or notice series of strange coincidences without particular consequence, such as meeting people with the same last name on the same day, seeing several times the same combination of digits in unrelated situations, etc.

An Austrian biologist dr. Paul Kammerer (1880-1926) was the first scientist to study this law. Although his book [Km] has attracted a lot of attention with its numerous suggestive examples, the scientific value of his “statistical” interpretation is rather questionable. Kammerer himself lost authority due to accusations of manipulating his (unrelated to our topic) biological experiments.

Also some very serious scientists such as Swiss professor of philosophy Karl Gustav Jung (1875-1961), and a Nobel prize winner in physics, Austrian, Wolfgang Pauli (1900-1958), fascinated by examples of “meaningful coincidences” conjectured the existence of undiscovered and mysterious “attracting” forces driving objects that are alike, or have common features, closer together in time and space, for which they coined a term “synchronicity”. This includes attracting of repetitions of rare events in time, i.e., the *law of series*. Critics of synchronicity claim that all such “unbelievable coincidences” and “series” occur at the rate complying with the statistics of pure randomness (see e.g. [Mi]). Human memory is keen to register them as more frequent simply because they are more distinctive.

To be precise, let us agree that an event repeats in time by “pure chance” when it follows a Poisson process. In a typical realization of such a process, the distribution of signals along the time axis reveals a natural tendency to create spontaneous clusters, which can be easily taken for series, but are in fact just a feature of the random (unbiased) behavior. In order to say that some signal process obeys the *law of series*, one should detect in this process a tendency to create clusters stronger than in the Poisson process. It is possible to formally define such tendency without referring to the multidimensional distributions of the

for the attracting to take effect. But the theory may apply to some rare events in computer sciences, genetics or in other areas.

RIGOROUS DEFINITIONS AND STATEMENTS

We establish the notation necessary to formulate the main results. Let $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ be an ergodic process on finitely many symbols, i.e., $\#\mathcal{P} < \infty$, σ is the standard left shift map and μ is an ergodic shift-invariant probability measure on $\mathcal{P}^{\mathbb{Z}}$. Most of the time, we will identify finite blocks with their cylinder sets, i.e., we agree that $\mathcal{P}^n = \bigvee_{i=0}^{n-1} \sigma^{-i}(\mathcal{P})$. Depending on the context, a block $B \in \mathcal{P}^n$ is attached to some coordinates or it represents a “word” which may appear in different places along the \mathcal{P} -names. We will also use the probabilistic language of random variables. Then $\mu\{R \in A\}$ ($A \subset \mathbb{R}$) will abbreviate $\mu(\{x \in \mathcal{P}^{\mathbb{Z}} : R(x) \in A\})$. Recall, that if the random variable R is nonnegative and $F(t) = \mu\{R \leq t\}$ is its distribution function, then the expected value of R equals $\int_0^\infty 1 - F(t) dt$.

For a set B of positive measure let R_B and \bar{R}_B denote the random variables defined on B (with the conditional measure $\mu_B = \frac{\mu}{\mu(B)}$) as the absolute and normalized first return time to B , respectively, i.e.,

$$R_B(y) = \min\{i > 0, \sigma^i(y) \in B\}, \quad \bar{R}_B(y) = \mu(B)R_B(y).$$

We denote by $\tilde{F}_B(t)$ the distribution function of \bar{R}_B . Notice that, by the Kac Theorem ([Kc]), the expected value of R_B equals $\frac{1}{\mu(B)}$, hence that of \bar{R}_B is 1 (that is why we call it “normalized”). We also define

$$G_B(t) = \int_0^t 1 - \tilde{F}_B(s) ds.$$

Clearly, $G_B(t) \leq \min\{t, 1\}$ and the equality holds when $\tilde{F}_B(t) = 1_{[1, \infty)}$, that is, when B occurs precisely with equal gaps (i.e., periodically); the gap size then equals $\frac{1}{\mu(B)}$.

Similarly, let V_B be the random variable defined on $\mathcal{P}^{\mathbb{Z}}$ as the *hitting time statistic*, i.e., the waiting time for the first visit in B (the defining formula is the same as for R_B , but this time it is regarded on the whole space with the measure μ). Further, let $\bar{V}_B = \mu(B)V_B$, called, by analogy, *the normalized hitting time* (although the expected value of this variable need not be equal to 1). By ergodicity, V_B and \bar{V}_B are well defined. By an elementary consideration of the skyscraper above B , one easily verifies, that the distribution function F_B of \bar{V}_B satisfies, for every $t \geq 0$, the inequalities:

$$G_B(t) - \mu(B) \leq F_B(t) \leq G_B(t)$$

(see [H-L-V] for more details). Because we deal with long blocks (so that, by the Shannon-McMillan-Breiman Theorem, $\mu(B)$ is, with high probability, very small), we will often replace F_B by G_B .

The key notions of this work are defined below:

Definition 1. We say that the visits to B *attract* (resp. *repel*) each other with intensity ϵ from a distance $t > 0$, if

$$F_B(t) \leq 1 - e^{-t} - \epsilon \quad (\text{resp. if } F_B(t) \geq 1 - e^{-t} + \epsilon).$$

We abbreviate that B *attracts* (*repels*) with intensity ϵ if its visits attract (repel) each other with intensity ϵ from some distance t .

Definition 2. We say that a process has *unbiased behavior* if there exist collections $\mathcal{B}_n \subset \mathcal{P}^n$ satisfying $\mu(\bigcup \mathcal{B}_n) \rightarrow 1$, such that $F_{B_n}(t) \rightarrow 1 - e^{-t}$ pointwise as $n \rightarrow \infty$, for any sequence of blocks $B_n \in \mathcal{B}_n$.

Definition 3. We say that a process reveals *strong attracting*, if there is a subset $\mathbb{N}' \subset \mathbb{N}$ of upper density 1, and collections $\mathcal{B}_{n'} \in \mathcal{P}^{n'}$ for $n' \in \mathbb{N}'$, satisfying $\mu(\bigcup \mathcal{B}_{n'}) \rightarrow 1$, such that $F_{B_{n'}}(t) \rightarrow 0$ pointwise as $n' \rightarrow \infty$, for any sequence of blocks $B_{n'} \in \mathcal{B}_{n'}$.

Let us explain why we use the terms “attracting” and “repelling”. We will compare $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ with an independent Bernoulli process which is unbiased, i.e., for any long block B , $\tilde{F}_B(t) \approx 1 - e^{-t}$ (and also $F_B(t) \approx 1 - e^{-t}$) with high uniform accuracy (much better than ϵ). Fix some $t > 0$. Consider the random variable I counting the number of occurrences of B in the time period $[0, \frac{t}{\mu(B)}]$. The expected value of I equals $\mu(B) \lfloor \frac{t}{\mu(B)} \rfloor \approx t$ (up to the ignorable error $\mu(B)$). On the other hand, $\mu\{I > 0\} = \mu\{V_B \leq \frac{t}{\mu(B)}\} = F_B(t)$. The ratio $\frac{t}{F_B(t)}$ represents the conditional expected value of I on the set $\{I > 0\}$, i.e., the expected number of occurrences of B in all intervals with at least one occurrence. Attracting from the distance t means that $F_B(t)$ is smaller (by ϵ) in $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ than in an independent Bernoulli process, i.e., that the above conditional expected value is larger in $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ than in the independent process. This fact can be further expressed as follows: If we observe the process $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ for time $\frac{t}{\mu(B)}$ (which is our “memory length” or “lifetime of the observer”) and we happen to see the event B during this time at least once, then the expected number of times we will observe the event B is larger than the analogous value for a cylinder of the same measure in the independent Bernoulli process. The first occurrence of B “attracts” its further repetitions. The interpretation of repelling is symmetric.

Obviously, occurrences of an event may simultaneously repel from one distance and attract from another. Notice, that the maximal intensity of repelling is e^{-1} achieved at $t = 1$ when B appears periodically (this implies repelling from all distances). The intensity of attracting can be arbitrarily close to 1, which happens when $F_B(t)$ (hence also $G_B(t)$) remains near zero for some large t (in particular this implies attracting from nearly all distances, except very small and very large ones, where marginal repelling can occur). It is easy to see that such case happens exactly when the distribution of the normalized return time is nearly concentrated at zero, i.e., when most points in the set B return after a time considerably smaller than $\frac{1}{\mu(B)}$. Because the expected value of the return time equals $\frac{1}{\mu(B)}$, there must be a small portion of B with extremely large values of the return time. In such case the event B appears in long series of high frequency, compensated by huge gaps of nearly complete absence. This is the essence of our notion of strong attracting.

The first main result follows:

Theorem 1. *If $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ is ergodic and has positive entropy, then for every $\epsilon > 0$ the measure of the union of all n -blocks $B \in \mathcal{P}^n$ which repel with intensity ϵ , converges to zero as n grows to infinity.*

Obviously, Theorem 1 does not exclude the unbiased behavior. For example, a Bernoulli process with the independent generator is unbiased. In fact, it follows from the results of [A-G], [H-S-V], that any process with a sufficient rate of mixing is unbiased (unbiased behavior is implied by “exponential asymptotics”). Nevertheless, our second theorem will say in particular, that processes with the unbiased behavior are extremely exceptional among positive entropy processes.

Let (X, μ) be a standard probability space, and let m denote either a finite integer or the countable cardinal \aleph_0 . The Rokhlin metric endows the collection of all measurable μ -distinguishable partitions \mathcal{P} of X into at most m elements with a topology of a Polish space.

Theorem 2. *Let (X, μ, T) be an ergodic measure-preserving transformation of a standard probability space, with positive entropy. Fix some $2 \leq m \leq \aleph_0$. Then, in the Polish space of all measurable partitions \mathcal{P} of X into at most m elements, there is a dense G_δ subset such that every partition in this subset generates a process which reveals strong attracting.*

Because partitions generating positive entropy form a dense open set (see Fact 5 below), we obtain that in a positive entropy measure preserving system a typical partition has both positive entropy and strong attracting.

MORE NOTATION AND PRELIMINARY FACTS

We now establish further notation and preliminaries needed in the proofs. If $\mathbb{A} \subset \mathbb{Z}$ then we will write $\mathcal{P}^{\mathbb{A}}$ to denote the partition or sigma-field $\bigvee_{i \in \mathbb{A}} \sigma^{-i}(\mathcal{P})$. We will abbreviate $\mathcal{P}^n = \mathcal{P}^{[0, n]}$, $\mathcal{P}^{-n} = \mathcal{P}^{[-n, -1]}$, $\mathcal{P}^- = \mathcal{P}^{(-\infty, -1]}$ (a “finite future”, a “finite past”, and the “full past” of the process).

We assume familiarity of the reader with the basics of entropy for finite partitions and sigma-fields in a standard probability space. Our notation is compatible with [P] and we refer the reader to this book, as well as to [Sh] and [Wa], for background and proofs. In particular, we will be using the following:

- * The entropy of a partition equals $H(\mathcal{P}) = -\sum_{A \in \mathcal{P}} \mu(A) \log_2(\mu(A))$.
- * For two finite partitions \mathcal{P} and \mathcal{B} , the conditional entropy $H(\mathcal{P}|\mathcal{B})$ is equal to $\sum_{B \in \mathcal{B}} \mu(B) H_B(\mathcal{P})$, where H_B is the entropy evaluated for the conditional measure μ_B on B .
- * The same formula holds for conditional entropy given a sub-sigma-field \mathcal{C} , i.e.,

$$\sum_{B \in \mathcal{B}} \mu(B) H_B(\mathcal{P}|\mathcal{C}) = H(\mathcal{P}|\mathcal{B} \vee \mathcal{C}).$$

- * The entropy of the process is given by any one of the formulas below

$$h = H(\mathcal{P}|\mathcal{P}^-) = \frac{1}{r} H(\mathcal{P}^r|\mathcal{P}^-) = \lim_{r \rightarrow \infty} \frac{1}{r} H(\mathcal{P}^r).$$

We will exploit the notion of ϵ -independence for partitions and sigma-fields. The definition below is an adaptation from [Sh], where it concerns finite partitions only. See also [Sm] for treatment of countable partitions. Because “ ϵ ” is reserved for the intensity of repelling, we will speak about β -independence.

Definition 4. Fix $\beta > 0$. A partition \mathcal{P} is said to be β -independent of a sigma-field \mathcal{B} if for any \mathcal{B} -measurable countable partition \mathcal{B}' holds

$$\sum_{A \in \mathcal{P}, B \in \mathcal{B}'} |\mu(A \cap B) - \mu(A)\mu(B)| \leq \beta.$$

A process $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ is called a β -independent process if \mathcal{P} is β -independent of the past \mathcal{P}^- .

A partition \mathcal{P} is independent of another partition or a sigma-field \mathcal{B} if and only if $H(\mathcal{P}|\mathcal{B}) = H(\mathcal{P})$. The following approximate version of this fact holds (see [Sh, Lemma 7.3] for finite partitions, from which the case of a sigma-field is easily derived).

Fact 1. A partition \mathcal{P} is β -independent of another partition or a sigma-field \mathcal{B} if $H(\mathcal{P}|\mathcal{B}) \geq H(\mathcal{P}) - \xi$, for ξ sufficiently small. \square

In course of the proof, a certain lengthy condition will be in frequent use. Let us introduce an abbreviation:

Definition 5. Given a partition \mathcal{P} of a space with a probability measure μ and $\delta > 0$, we will say that a property $\Phi(A)$ holds for $A \in \mathcal{P}$ with μ -tolerance δ if

$$\mu\left(\bigcup\{A \in \mathcal{P} : \Phi(A)\}\right) \geq 1 - \delta.$$

We shall also need an elementary estimate, whose proof is an easy exercise.

Fact 2. For each $A \in \mathcal{P}$, $H(\mathcal{P}) \leq (1 - \mu(A)) \log_2(\#\mathcal{P}) + 1$. \square

In addition to the random variables of the absolute and normalized return times R_B and \bar{R}_B , we will also use the analogous notions of the k^{th} absolute return time

$$R_B^{(k)} = \min\{i : \#\{0 < j \leq i : \sigma^j(y) \in B\} = k\},$$

and of the normalized k^{th} return time $\bar{R}_B^{(k)} = \mu(B)R_B^{(k)}$ (both defined on B), with $\tilde{F}_B^{(k)}$ always denoting the distribution function of the latter. Clearly, the expected value of $\bar{R}_B^{(k)}$ equals k .

THE IDEA OF THE PROOF AND THE BASIC LEMMA

Before we pass to the formal proof of Theorem 1, we would like to have the reader oriented in the mainframe of the idea behind it. We intend to estimate (from above, by $1 - e^{-t} + \epsilon$) the function G_{BA} (replacing F_{BA}), for long blocks of the form $BA \in \mathcal{P}^{[-n,r]}$. The “positive” part A has a fixed length r , while we allow the “negative” part B to be arbitrarily long. There are two key ingredients leading to the estimation. The first one, contained in Lemma 3, is the observation that for a fixed typical $B \in \mathcal{P}^{-n}$, the part of the process induced on B (with the conditional measure μ_B) generated by the partition \mathcal{P}^r , is not only a β -independent process, but it is also β -independent of many returns times $R_B^{(k)}$ of the cylinder B (see the Figure 2).

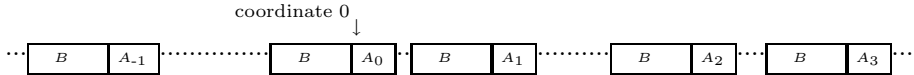


Figure 2: The process $\dots A_{-1}A_0A_1A_2\dots$ of r -blocks following the copies of B is a β -independent process with additional β -independence properties of the positioning of the copies of B .

This allows us to decompose (with high accuracy) the distribution function \tilde{F}_{BA} of the normalized return time of BA as follows:

$$\begin{aligned} \tilde{F}_{BA}(t) &= \mu_{BA}\{\bar{R}_{BA} \leq t\} = \mu_{BA}\{R_{BA} \leq \frac{t}{\mu(BA)}\} = \\ \sum_{k \geq 1} \mu_{BA}\{R_A^{(B)} = k, R_B^{(k)} \leq \frac{t}{p\mu(B)}\} &\approx \sum_{k \geq 1} \mu_{BA}\{R_A^{(B)} = k\} \cdot \mu_B\{\bar{R}_B^{(k)} \leq \frac{t}{p}\} \approx \\ &\sum_{k \geq 1} p(1-p)^{k-1} \cdot \tilde{F}_B^{(k)}\left(\frac{t}{p}\right), \end{aligned}$$

where $R_A^{(B)}$ denotes the first (absolute) return time of A in the process induced on B , and $p = \mu_B(A)$.

The second key observation is, assuming for simplicity full independence, that when trying to model some repelling for the blocks BA , we ascertain that it is largest, when the occurrences of B are purely periodic. Any deviation from periodicity of the B 's may only lead to increasing the intensity of attracting between the copies of BA , never that of repelling. We will explain this phenomenon more formally in a moment. Now, if B does appear periodically, then the normalized return time of BA is governed by the same geometric distribution as the normalized return time of A in the independent process induced on B . If p is small, this geometric distribution function becomes nearly the unbiased exponential law $1 - e^{-t}$. The smallness of p is *a priori* regulated by the choice of the parameter r (Lemma 1).

The phenomena that, assuming full independence, the repelling of BA is maximized by periodic occurrences of B , and that even then there is nearly no repelling, is captured by the following elementary lemma, which will be also useful later, near the end of the rigorous proof.

Lemma 0. *Fix some $p \in (0, 1)$. Let $\tilde{F}^{(k)}$ ($k \geq 1$) be a sequence of distribution functions on $[0, \infty)$ such that the expected value of the distribution associated to $\tilde{F}^{(k)}$ equals k . Define*

$$\tilde{F}(t) = \sum_{k \geq 1} p(1-p)^{k-1} \tilde{F}^{(k)}\left(\frac{t}{p}\right), \quad \text{and} \quad G(t) = \int_0^t 1 - \tilde{F}(s) ds.$$

Then $G(t) \leq \frac{1}{\log e_p} (1 - e_p^{-t})$, where $e_p = (1-p)^{-\frac{1}{p}}$.

Proof. We have

$$G(t) = \sum_{k \geq 1} p(1-p)^{k-1} \int_0^t 1 - \tilde{F}^{(k)}\left(\frac{s}{p}\right) ds.$$

We know that $\tilde{F}^{(k)}(t) \in [0, 1]$ and that $\int_0^\infty 1 - \tilde{F}^{(k)}(s) ds = k$ (the expected value). With such constraints, it is the indicator function $1_{[k, \infty)}$ that maximizes the integrals from 0 to t simultaneously for every t (because the “mass” k above the graph is, for such choice of the function $\tilde{F}^{(k)}$, swept maximally to the left). The rest follows by direct calculations:

$$\begin{aligned} G(t) &\leq \sum_{k \geq 1} p(1-p)^{k-1} \int_0^t 1_{[0, k)}\left(\frac{s}{p}\right) ds = \int_0^t \sum_{k=\lceil \frac{s}{p} \rceil}^\infty p(1-p)^{k-1} ds = \\ &\int_0^t (1-p)^{\lceil \frac{s}{p} \rceil} ds \leq \frac{(1-p)^{\frac{t}{p}} - 1}{\log(1-p)^{\frac{1}{p}}}. \quad \square \end{aligned}$$

Recall that the maximizing distribution functions $\tilde{F}_B^{(k)} = 1_{[k, \infty)}$ occur, for the normalized return time of a set B , precisely when B is visited periodically. This explains our former statement on this subject.

Let us comment a bit more on the first key ingredient, the β -independence. Establishing it is the most complicated part of the argument. The idea is to prove conditional (given a “finite past” \mathcal{P}^{-n}) β -independence of the “present” \mathcal{P}^r from jointly the full past and a large part of the future, responsible for the return times

of majority of the blocks $B \in \mathcal{P}^{-n}$. But the future part must not be too large. Let us mention the existence of “bilaterally deterministic” processes with positive entropy (first discovered by Gurevič [G], see also [O-W1]), in which the sigma-fields generated by the coordinates $(-\infty, -m] \cup [m, \infty)$ do not decrease with m to the Pinsker factor; they are all equal to the entire sigma-field. (Coincidentally, our Example 1 has precisely this property; see the Remark 2.) Thus, in order to maintain any trace of independence of the “present” from our sigma-field already containing the entire past, its part in the future must be selected with an extreme care. Let us also remark that an attempt to save on the future sigma-fields by adjusting them individually to each block $B_0 \in \mathcal{P}^{-n}$ falls short, mainly because of the “off diagonal effect”; suppose \mathcal{P}^r is conditionally (given \mathcal{P}^{-n}) nearly independent of a sigma-field which determines the return times of only one selected block $B_0 \in \mathcal{P}^{-n}$. The independence still holds conditionally given any cylinder $B \in \mathcal{P}^{-n}$ from a collection of a large measure, but unfortunately, this collection can always miss the selected cylinder B_0 . In Lemmas 2 and 3, we succeed in finding a sigma-field (containing the full past and a part of the future), of which \mathcal{P}^r is conditionally β -independent, and which “nearly determines”, for majority of blocks $B \in \mathcal{P}^{-n}$, some finite number of their sequential return times (probably not all of them). This finite number is sufficient to allow the described earlier decomposition of the distribution function \tilde{F}_{BA} .

THE PROOF OF THEOREM 1

Throughout the sequel we assume ergodicity and that the entropy h of $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ is positive. We begin our computations with an auxiliary lemma allowing us to assume (by replacing \mathcal{P} by some \mathcal{P}^r) that the elements of the “present” partition are small, relatively in most of $B \in \mathcal{P}^n$ and for every n . Note that the Shannon-McMillan-Breiman Theorem is insufficient: for the conditional measure the error term in that theorem depends increasingly on n , which we do not fix.

Lemma 1. *For each δ there exists an $r \in \mathbb{N}$ such that for every $n \in \mathbb{N}$ the following holds for $B \in \mathcal{P}^{-n}$ with μ -tolerance δ :*

$$\text{for every } A \in \mathcal{P}^r, \quad \mu_B(A) \leq \delta.$$

Proof. Let α be so small that

$$\sqrt{\alpha} \leq \delta \quad \text{and} \quad \frac{h - 3\sqrt{\alpha}}{h + \alpha} \geq 1 - \frac{\delta}{2},$$

and set $\gamma = \frac{\alpha}{\log_2(\#\mathcal{P})}$. Let r be so big that

$$\frac{1}{r} \leq \alpha, \quad \frac{1}{r(h + \alpha)} \leq \frac{\delta}{2},$$

and that there exists a collection $\overline{\mathcal{P}^r}$ of no more than $2^{r(h+\alpha)} - 1$ elements of \mathcal{P}^r whose joint measure μ exceeds $1 - \gamma$ (by the Shannon-McMillan-Breiman Theorem).

Let $\widetilde{\mathcal{P}^r}$ denote the partition into the elements of $\overline{\mathcal{P}^r}$ and the complement of their union, and let \mathcal{R} be the partition into the remaining elements of \mathcal{P}^r and the complement of their union, so that $\mathcal{P}^r = \widetilde{\mathcal{P}^r} \vee \mathcal{R}$. For any n we have

$$\begin{aligned} rh &= H(\mathcal{P}^r | \mathcal{P}^{-n}) \leq H(\mathcal{P}^r | \mathcal{P}^{-n}) = H(\widetilde{\mathcal{P}^r} \vee \mathcal{R} | \mathcal{P}^{-n}) = \\ &= H(\widetilde{\mathcal{P}^r} | \mathcal{R} \vee \mathcal{P}^{-n}) + H(\mathcal{R} | \mathcal{P}^{-n}) \leq H(\widetilde{\mathcal{P}^r} | \mathcal{P}^{-n}) + H(\mathcal{R}) \leq \\ &\quad \sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\widetilde{\mathcal{P}^r}) + \gamma r \log_2(\#\mathcal{P}) + 1 \end{aligned}$$

(we have used Fact 2 for the last passage). After dividing by r , we obtain

$$\sum_{B \in \mathcal{P}^{-n}} \mu(B) \frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \geq h - \gamma \log_2(\#\mathcal{P}) - \frac{1}{r} \geq h - 2\alpha.$$

Because each term $\frac{1}{r} H_B(\widetilde{\mathcal{P}^r})$ is not larger than $\frac{1}{r} \log_2(\#\widetilde{\mathcal{P}^r})$ which was set to be at most $h + \alpha$, we deduce that

$$\frac{1}{r} H_B(\widetilde{\mathcal{P}^r}) \geq h - 3\sqrt{\alpha}$$

holds for $B \in \mathcal{P}^{-n}$ with μ -tolerance $\sqrt{\alpha}$, hence also with μ -tolerance δ . On the other hand, by Fact 2, for any B and $A \in \widetilde{\mathcal{P}^r}$, holds:

$$H_B(\widetilde{\mathcal{P}^r}) \leq (1 - \mu_B(A)) \log_2(\#\widetilde{\mathcal{P}^r}) + 1 \leq (1 - \mu_B(A))r(h + \alpha) + 1.$$

Combining the last two displayed inequalities we establish that, with μ -tolerance δ for $B \in \mathcal{P}^{-n}$ and then for every $A \in \widetilde{\mathcal{P}^r}$, holds

$$1 - \mu_B(A) \geq \frac{h - 3\sqrt{\alpha}}{h + \alpha} - \frac{1}{r(h + \alpha)} \geq 1 - \delta.$$

So, $\mu_B(A) \leq \delta$. Because \mathcal{P}^r refines $\widetilde{\mathcal{P}^r}$, the elements of \mathcal{P}^r are also not larger than δ . \square

We continue the proof with a lemma which can be deduced from [Ru, Lemma 3]. We provide a direct proof. For $\alpha > 0$ and $M \in \mathbb{N}$ let

$$S(M, \alpha) = \bigcup_{m \in \mathbb{Z}} [mM + \alpha M, (m+1)M - \alpha M) \cap \mathbb{Z}.$$

Lemma 2. *For fixed α and r there exists M_0 such that for every $M \geq M_0$ holds,*

$$H(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) \geq rh - \alpha$$

(see the Figure 3).

*****○●.....*****.....*****.....*****.....

Figure 3. The circles indicate the coordinates 0 through $r-1$, the conditioning sigma-field is over the coordinates marked by stars, which includes the entire past and part of the future with gaps of size $2\alpha M$ repeated periodically with period M (the first gap is half the size).

Proof. First assume that $r = 1$. Denote also

$$S'(M, \alpha) = \bigcup_{m \in \mathbb{Z}} [mM + \alpha M, (m+1)M) \cap \mathbb{Z}.$$

Let M be so large that $H(\mathcal{P}^{(1-\alpha)M}) < (1-\alpha)M(h + \gamma)$, where $\gamma = \frac{\alpha^2}{2(1-\alpha)}$. Then, for any $m \geq 1$,

$$H(\mathcal{P}^{S'(M, \alpha) \cap [0, mM]} | \mathcal{P}^-) \leq H(\mathcal{P}^{S'(M, \alpha) \cap [0, mM]}) < (1-\alpha)mM(h + \gamma).$$

Because $H(\mathcal{P}^{[0,mM]}|\mathcal{P}^-) = mMh$, the complementary part of entropy must exceed $mMh - (1 - \alpha)mM(h + \gamma)$ (which equals $\alpha mM(h - \frac{\alpha}{2})$), i.e., we have

$$H(\mathcal{P}^{[0,mM] \setminus S'(M,\alpha)}|\mathcal{P}^- \vee \mathcal{P}^{S'(M,\alpha) \cap [0,mM]}) > \alpha mM(h - \frac{\alpha}{2}).$$

Breaking the last entropy term as a sum over $j \in [0, mM] \setminus S'(M, \alpha)$ of the conditional entropies of $\sigma^{-j}(\mathcal{P})$ given the sigma-field over all coordinates left of j and all coordinates from $S'(M, \alpha) \cap [0, mM)$ right of j , and because every such term is at most h , we deduce that more than half of these terms reach or exceed $h - \alpha$. So, a term not smaller than $h - \alpha$ occurs for a j within one of the gaps in the left half of $[0, mM)$. Shifting by j , we obtain $H(\mathcal{P}|\mathcal{P}^- \vee \sigma^i(\mathcal{P}^{S'(M,\alpha) \cap [0, \frac{mM}{2}]}) \geq h - \alpha$, where $i \in [0, \alpha M)$ denotes the relative position of j in the gap. As we increase m , one value i will repeat in this role along a subsequence m' . The operation \vee is continuous for increasing sequences of sigma-fields, hence $\mathcal{P}^- \vee \sigma^i(\mathcal{P}^{S'(M,\alpha) \cap [0, \frac{m'M}{2}]})$ converges over m' to $\mathcal{P}^- \vee \sigma^i(\mathcal{P}^{S'(M,\alpha)})$. The entropy is continuous for such passage, hence $H(\mathcal{P}|\mathcal{P}^- \vee \sigma^i(\mathcal{P}^{S'(M,\alpha)})) \geq h - \alpha$. The assertion now follows because $S(M, \alpha)$ is contained in $S'(M, \alpha)$ shifted to the left by any $i \in [0, \alpha M)$.

Finally, if $r > 1$, we can simply argue for \mathcal{P}^r replacing \mathcal{P} . This will impose that M_0 and M are divisible by r , but it is not hard to see that for large M the argument works without divisibility at a cost of a slight adjustment of α . \square

For a long block $B \in \mathcal{P}^{-n}$ let $((\mathcal{P}_B^r)^\mathbb{Z}, \mu_B, \sigma_B)$ denote the process induced on B generated by the restriction \mathcal{P}_B^r of \mathcal{P}^r to B (σ_B is the first return time map on B). The following lemma is the crucial item in our argument.

Lemma 3. *For every $\beta > 0$, $r \in \mathbb{N}$ and $K \in \mathbb{N}$ there exists n_0 such that for every $n \geq n_0$, with μ -tolerance β for $B \in \mathcal{P}^{-n}$, with respect to μ_B , \mathcal{P}^r is β -independent of jointly the past \mathcal{P}^- and the first K return times to B , $R_B^{(k)}$ ($k \in [1, K]$). In particular, $((\mathcal{P}_B^r)^\mathbb{Z}, \mu_B, \sigma_B)$ is a β -independent process.*

Proof. We choose ξ according to Fact 1, so that $\frac{\beta}{2}$ -independence is implied. Let α satisfy

$$0 < \frac{2\alpha}{h-\alpha} < 1, \quad 18K\sqrt{\alpha} < 1, \quad \sqrt{2\alpha} < \xi, \quad K\sqrt[4]{\alpha} < \frac{\beta}{2}.$$

Let n_0 be so large that $H(\mathcal{P}^r|\mathcal{P}^{-n}) < rh + \alpha$ for every $n \geq n_0$ and that for every $k \in [1, K]$ with μ -tolerance α for $B \in \mathcal{P}^{-n}$ holds

$$\mu_B\{2^{n(h-\alpha)} \leq R_B^{(k)} \leq 2^{n(h+\alpha)}\} > 1 - \alpha$$

(we are using Ornstein-Weiss Theorem [O-W2]; the multiplication by k , which should appear for the k^{th} return time, is consumed by α in the exponent). Let $M_0 \geq 2^{n_0(h-\alpha)}$ be so large that the assertion of Lemma 2 holds for α , r and M_0 , and that for every $M \geq M_0$,

$$(M+1)^{1+\frac{2\alpha}{h-\alpha}} < \alpha M^2 \quad \text{and} \quad \frac{\log_2(M+1)}{M(h-\alpha)} < \alpha.$$

We can now redefine (enlarge) n_0 and M_0 so that $M_0 = \lfloor 2^{n_0(h-\alpha)} \rfloor$. Similarly, for each $n \geq n_0$ we set $M_n = \lfloor 2^{n(h-\alpha)} \rfloor$. Observe, that the interval where the first K returns of most n -blocks B may occur (up to probability α), is contained in $[M_n, \alpha M_n^2]$ (because $2^{n(h+\alpha)} \leq (M_n + 1)^{1+\frac{2\alpha}{h-\alpha}} < \alpha M_n^2$).

At this point we fix some $n \geq n_0$. The idea is to carefully select an M between M_n and $2M_n$ (hence not smaller than M_0), such that the initial K returns of nearly

every n -block happen most likely inside (with all its n symbols) the set $S(M, \alpha)$, so that they are “controlled” by the sigma-field $\mathcal{P}^{S(M, \alpha)}$. Let $\alpha' = \alpha + \frac{n}{M_n}$, so that every n -block overlapping with $S(M, \alpha')$ is completely covered by $S(M, \alpha)$. By the second assumption on $M \geq M_0$ and by the formula connecting M_n and n , we have $\alpha' < 2\alpha$. To define M we will invoke the triple Fubini Theorem. Fix $k \in [1, K]$ and consider the probability space

$$\mathcal{P}^{-n} \times [M_n, 2M_n] \times \mathbb{N}$$

equipped with the (discrete) measure \mathcal{M} whose marginal on $\mathcal{P}^{-n} \times [M_n, 2M_n]$ is the product of μ (more precisely, of its projection onto \mathcal{P}^{-n}) with the uniform distribution on the integers in $[M_n, 2M_n]$, while, for fixed B and M , the measure on the corresponding \mathbb{N} -section is the distribution of the random variable $R_B^{(k)}$. In this space let S be the set whose \mathbb{N} -section for a fixed M (and any fixed B) is the set $S(M, \alpha')$. We claim that for every $l \in [M_n, \alpha M_n^2] \cap \mathbb{N}$ (and any fixed B) the $[M_n, 2M_n]$ -section of S has measure exceeding $1 - 16\alpha$. This is quite obvious (even for every $l \in [M_n, \infty)$ and with $1 - 15\alpha$) if $[M_n, 2M_n]$ is equipped with the normalized Lebesgue measure (see the Figure 4).

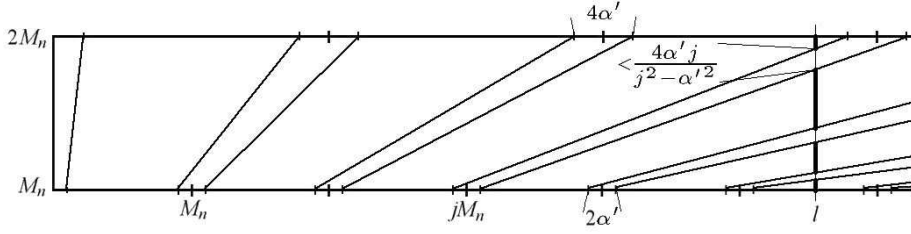


Figure 4: The complement of S splits into thin skew strips shown in the picture. The normalized Lebesgue measure of any vertical section of the j^{th} strip (starting at jM_n with $j \geq 1$) is at most $\frac{4\alpha'j}{j^2 - \alpha'^2} \leq \frac{5\alpha'}{j} \leq \frac{10\alpha}{j}$. Each vertical line at $l \geq M_n$ intersects strips with indices $j, j+1, j+2$ up to at most $2j$ (for some j), so the joint measure of the complement of the section of S does not exceed 15α .

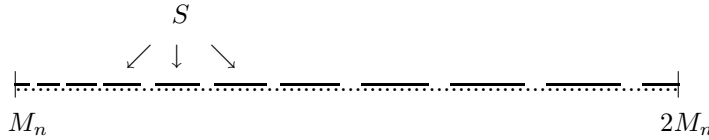


Figure 5: The discretization replaces the Lebesgue measure by the uniform measure on M_n integers, thus the measure of any interval can deviate from its Lebesgue measure by at most $\frac{1}{M_n}$. For $l \leq \alpha M_n^2$ the corresponding section of S (in this picture drawn horizontally) consists of at most αM_n intervals, so its measure can deviate by no more than α .

In the discrete case, however, *a priori* it might happen that the integers along some $[M_n, 2M_n]$ -section often “miss” the section of S leading to a decreased measure value. (For example, it is easy to see that for $l = (2M_n)!$ the measure of the section of S is zero.) But because we restrict to $l \leq \alpha M_n^2$, the discretization does not affect the measure of the section of S by more than α , and the estimate with $1 - 16\alpha$ holds (see the Figure 5 above).

Taking into account all other inaccuracies (the smaller than α part of S outside $[M_n, \alpha M_n^2]$ and the smaller than α part of S projecting onto blocks B which do not obey the Ornstein-Weiss return time estimate) it is safe to claim that

$$\mathcal{M}(S) > 1 - 18\alpha.$$

This implies that for every M from a set of measure at least $1 - 18\sqrt{\alpha}$ the measure of the $(\mathcal{P}^{-n} \times \mathbb{N})$ -section of S is larger than or equal to $1 - \sqrt{\alpha}$. For every such M , with μ -tolerance $\sqrt[4]{\alpha}$ for $B \in \mathcal{P}^{-n}$, the probability μ_B that the k^{th} repetition of B falls in $S(M, \alpha')$ (hence with all its n terms inside the set $S(M, \alpha)$) is at least $1 - \sqrt[4]{\alpha}$.

Because $18K\sqrt{\alpha} < 1$, there exists at least one M for which the above holds for every $k \in [1, K]$. This is our final choice of M which from now on remains fixed. For this M , and for cylinders B chosen with μ -tolerance $K\sqrt[4]{\alpha}$, each of the considered K returns of B with probability $1 - \sqrt[4]{\alpha}$ falls (with all its coordinates) inside $S(M, \alpha)$. Thus, for such a B , with probability $1 - K\sqrt[4]{\alpha}$ the same holds simultaneously for all K return times. In other words, there is a set U_B of measure not exceeding $K\sqrt[4]{\alpha}$ outside of which $R_B^{(k)} = \tilde{R}_B^{(k)}$, where $\tilde{R}_B^{(k)}$ is defined as the time of the k^{th} fully visible inside $S(M, \alpha)$ return of B . Notice that $\tilde{R}_B^{(k)}$ is $\mathcal{P}^{S(M, \alpha)}$ -measurable.

Let us go back to our entropy estimates. We have, by Lemma 2,

$$\begin{aligned} \sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) &= H(\mathcal{P}^r | \mathcal{P}^{-n} \vee \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) = \\ &H(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) \geq rh - \alpha \geq H(\mathcal{P}^r | \mathcal{P}^{-n}) - 2\alpha = \\ &\sum_{B \in \mathcal{P}^{-n}} \mu(B) H_B(\mathcal{P}^r) - 2\alpha. \end{aligned}$$

Because $H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) \leq H_B(\mathcal{P}^r)$ for every B , we deduce that with μ -tolerance $\sqrt{2\alpha}$ for $B \in \mathcal{P}^{-n}$ must hold

$$H_B(\mathcal{P}^r | \mathcal{P}^- \vee \mathcal{P}^{S(M, \alpha)}) \geq H_B(\mathcal{P}^r) - \sqrt{2\alpha} \geq H_B(\mathcal{P}^r) - \xi.$$

Combining this with the preceding arguments, with μ -tolerance $K\sqrt[4]{\alpha} + \sqrt{2\alpha} < \beta$ for $B \in \mathcal{P}^{-n}$ both the above entropy inequality holds, and we have the estimates of the measures of sets U_B . By the choice of ξ , we obtain that with respect to μ_B , \mathcal{P}^r is jointly $\frac{\beta}{2}$ -independent of the past and the modified return times $\tilde{R}_B^{(k)}$ ($k \in [1, K]$). Because $\mu(U_B) \leq K\sqrt[4]{\alpha} < \frac{\beta}{2}$, this clearly implies β -independence if each $\tilde{R}_B^{(k)}$ is replaced by $R_B^{(k)}$. \square

To complete the proof of Theorem 1 it now remains to put the items together.

Proof of Theorem 1. Fix an $\epsilon > 0$. On $[0, \infty)$, the functions

$$g_p(t) = \min\{1, \frac{1}{\log e_p}(1 - e_p^{-t}) + pt\},$$

where $e_p = (1 - p)^{-\frac{1}{p}}$, decrease uniformly to $1 - e^{-t}$ as $p \rightarrow 0^+$. So, let δ be such that $g_\delta(t) \leq 1 - e^{-t} + \epsilon$ for every t . We also assume that

$$(1 - 2\delta)(1 - \delta) \geq 1 - \epsilon.$$

Let r be specified by Lemma 1, so that $\mu_B(A) \leq \delta$ for every $n \geq 1$, every $A \in \mathcal{P}^r$ and for $B \in \mathcal{P}^{-n}$ with μ -tolerance δ . On the other hand, once r is fixed, the partition \mathcal{P}^r has at most $(\#\mathcal{P})^r$ elements, so with μ_B -tolerance δ for $A \in \mathcal{P}^r$, $\mu_B(A) \geq \delta(\#\mathcal{P})^{-r}$. Let \mathcal{A}_B be the subfamily of \mathcal{P}^r (depending on B) where this inequality holds. Let K be so large that for any $p \geq \delta(\#\mathcal{P})^{-r}$,

$$\sum_{k=K+1}^{\infty} p(1-p)^k < \frac{\delta}{2},$$

and choose $\beta < \delta$ so small that

$$(K^2 + K + 1)\beta < \frac{\delta}{2}.$$

The application of Lemma 3 now provides an n_0 such that for any $n \geq n_0$, with μ -tolerance β for $B \in \mathcal{P}^{-n}$, the process induced on B generated by \mathcal{P}^r has the desired β -independence properties involving the initial K return times of B . So, with tolerance $\delta + \beta < 2\delta$ we have both, the above β -independence and the estimate $\mu_B(A) < \delta$ for every $A \in \mathcal{P}^r$. Let \mathcal{B}_n be the subfamily of \mathcal{P}^{-n} where these two conditions hold. Fix some $n \geq n_0$.

Let us consider a cylinder set $B \cap A \in \mathcal{P}^{[-n, r]}$ (or, equivalently, the block BA), where $B \in \mathcal{B}_n$, $A \in \mathcal{A}_B$. The length of BA is $n + r$, which represents an arbitrary integer larger than $n_0 + r$. Notice that the family of such sets BA covers more than $(1 - 2\delta)(1 - \delta) \geq 1 - \epsilon$ of the space.

We will examine the distribution of the normalized first return time for BA . In addition to our customary notations of return times, let $R_A^{(B)}$ be the first (absolute) return time of A in $((\mathcal{P}_B^r)^{\mathbb{Z}}, \mu_B, \sigma_B)$, i.e., the variable defined on BA , counting the number of visits to B until the first return to BA . Let $p = \mu_B(A)$ (recall, this is not smaller than $\delta(\#\mathcal{P})^{-r}$). We have

$$\begin{aligned} \tilde{F}_{BA}(t) &= \mu_{BA}\{\overline{R}_{BA} \leq t\} = \mu_{BA}\{R_{BA} \leq \frac{t}{\mu(BA)}\} = \\ &= \sum_{k \geq 1} \mu_{BA}\{R_A^{(B)} = k, R_B^{(k)} \leq \frac{t}{p\mu(B)}\}. \end{aligned}$$

The k^{th} term of this sum equals

$$\frac{1}{p} \mu_B(\{A_k = A\} \cap \{A_{k-1} \neq A\} \cap \cdots \cap \{A_1 \neq A\} \cap \{A_0 = A\} \cap \{R_B^{(k)} \leq \frac{t}{p\mu(B)}\}),$$

where A_i is the r -block following the i^{th} copy of B (the counting starts from 0 at the copy of B positioned at $[-n, -1]$).

By Lemma 3, for $k \leq K$, in this intersection of sets each term is β -independent of the intersection right from it. So, proceeding from the left, we can replace the probabilities of the intersections by products of probabilities, allowing an error of β . Note that the last term equals $\mu_B\{\overline{R}_B^{(k)} \leq \frac{t}{p}\} = \tilde{F}_B^{(k)}(\frac{t}{p})$. Jointly, the inaccuracy will not exceed $(K + 1)\beta$:

$$\left| \mu_{BA}\{R_A^{(B)} = k, R_B^{(k)} \leq \frac{t}{p\mu(B)}\} - p(1-p)^{k-1} \tilde{F}_B^{(k)}(\frac{t}{p}) \right| \leq (K + 1)\beta.$$

Similarly, we also have $\left| \mu_{BA}\{R_A^{(B)} = k\} - p(1-p)^{k-1} \right| \leq K\beta$, hence the tail of the series $\mu_{BA}\{R_A^{(B)} = k\}$ above K is smaller than $K^2\beta$ plus the tail of the geometric series $p(1-p)^{k-1}$, which, by the fact that $p \geq \delta(\#\mathcal{P})^{-r}$, is smaller than $\frac{\delta}{2}$. Therefore

$$\tilde{F}_{BA}(t) \approx \sum_{k \geq 1} p(1-p)^{k-1} \tilde{F}_B^{(k)}(\frac{t}{p}),$$

up to $(K^2 + K + 1)\beta + \frac{\delta}{2} \leq \delta$, uniformly for every t . By the application of Lemma 0, G_{BA} satisfies

$$G_{BA}(t) \leq \min\{1, \frac{1}{\log e_p}(1 - e_p^{-t}) + \delta t\} \leq g_\delta(t) \leq 1 - e^t + \epsilon$$

(because $p \leq \delta$). We have proved that for our choice of ϵ and an arbitrary length $m \geq n_0 + r$, with μ -tolerance ϵ for the cylinders $C \in \mathcal{P}^m$, the intensity of repelling between visits to C is at most ϵ . This concludes the proof of Theorem 1. \square

PROOF OF THEOREM 2

This proof requires a number of technical ingredients, such as “semi-periodic markers” or short “transciently forbidden words”. The two facts below are standard exercises in ergodic theory and we only outline their proofs.

Fact 3. *In a process $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$ of positive entropy, where \mathcal{P} is finite or countable, for each $k \in \mathbb{N}$ and $\epsilon > 0$ there exist an $l \in \mathbb{N}$ and k words w_1, w_2, \dots, w_k of length l such that*

1. *each w_i starts and ends with the same symbol $a \in \mathcal{P}$, independent from i*
2. *each w_i has measure μ at most $\frac{\epsilon}{lk}$,*
3. *for each i the set*

$$w_i \setminus \bigcup_{j \neq i} \bigcup_{m=-l}^l \sigma^m(w_j)$$

has positive measure μ .

Proof. For 1. use recurrence in the k -fold product system, and for 2. use the Shannon-McMillan-Breiman Theorem. Condition 3. follows easily from the high complexity in positive entropy. \square

Fact 4. *In every measure-preserving system (X, μ, T) of positive entropy h , for each sufficiently large $r \in \mathbb{N}$ there exists a “semiperiodic r -marker”, i.e., a measurable set F such that the first return time R_F assumes only two values: r and $r + 1$.*

Proof. The system has a Bernoulli factor of entropy h . For large r the binary process obtained by random concatenations of two blocks, $0^{r-1}1$ and 0^r1 , is Bernoulli with entropy smaller than h , hence it is a factor of (X, μ, T) . The lift of the cylinder over 1 is the desired set F in X . \square

We are in a position to present the proof of Theorem 2.

Proof of Theorem 2. Fix $\epsilon > 0$, $t > 0$ and $N \in \mathbb{N}$. Consider the following property of a (finite or countable) partition \mathcal{P} : for every $n \in [N, N^2]$, $F_B(t) < \epsilon$ with μ -tolerance ϵ for $B \in \mathcal{P}^n$. (Recall that F_B denotes the distribution function of the normalized hitting time for B). It is easy to see that it holds on an open set $\mathcal{E}_{\epsilon, t, N}$ of partitions (both in the space of partitions into at most m elements and in the space of at most countable partitions); for each n we can take the same finite sets of “good” n -cylinders B for the partitions in a neighborhood of \mathcal{P} as for \mathcal{P} . Of course, the set

$$\mathcal{E}_{\epsilon, t} = \bigcup_{N \geq 1} \mathcal{E}_{\epsilon, t, N},$$

of partitions such that the same property holds for some N , is also open. The main effort in the proof will be to show that this set is also dense. Once this is done,

the proof is complete, because then the dense G_δ set of partitions which reveal strong attracting can be obtained by intersecting the sets $\mathcal{E}_{\epsilon,t}$ over countably many pairs (ϵ, t) with $\epsilon \rightarrow 0$ and $t \rightarrow \infty$. Notice that for any infinite sequence of natural numbers N the set $\bigcup[N, N^2]$ has upper density 1 in \mathbb{N} .

In order to prove the density of $\mathcal{E}_{\epsilon,t}$, fix a (finite or countable) partition \mathcal{P} . Set

$$k = \lceil \frac{2}{\epsilon} \rceil + 1, \quad \delta = \frac{1-\frac{\epsilon}{2}}{4k}, \quad M = 2kt.$$

Choose words w_1, w_2, \dots, w_k according to Fact 3. Let N be so large, that with μ -tolerance $\frac{\epsilon}{2}$ in every N -block, every word w_i occurs at least once so that it does not overlap with any other w_j (see condition 3. in Fact 3). Obviously, the same holds if N is replaced by any larger integer. For every $n \in [N, N^2]$ we can thus select a finite collection of “good” n -blocks which satisfy the above and cover $1 - \frac{\epsilon}{2}$ of the space. Let p be so large, that $\frac{2N^2}{p} < \frac{\epsilon}{2}$, and that every good n -block (for any $n \in [N, N^2]$) occurs at least M times in every, up to μ -tolerance δ , $\frac{p}{2}$ -block. Let $r = kp$.

Now we invoke the semiperiodic r -marker set F of Fact 4. Every \mathcal{P} -name can be divided at visits to F into a concatenation of r -blocks and $(r+1)$ -blocks. For simplicity, we will call all of them *component r -blocks*. Every component r -block C will be further decomposed as a concatenation of k p -blocks $C_1 C_2 \dots C_k$ (C_k is either a p -block or a $(p+1)$ -block, but again, for simplicity, we will call all these blocks p -blocks). We fix a symbol $b \neq a$ in \mathcal{P} (recall that a denotes the first and last symbol of each w_i). Now we modify the partition \mathcal{P} by changing the \mathcal{P} -names of points, as follows: In every \mathcal{P} -name we replace, for every i , every occurrence of w_i within every i^{th} p -block C_i of every component r -block C and within the first N^2 positions of the following p -block C_{i+1} (here $k+1=1$), by the word $w_0 = b^l$. Notice that there is no collision when overlapping words are replaced.

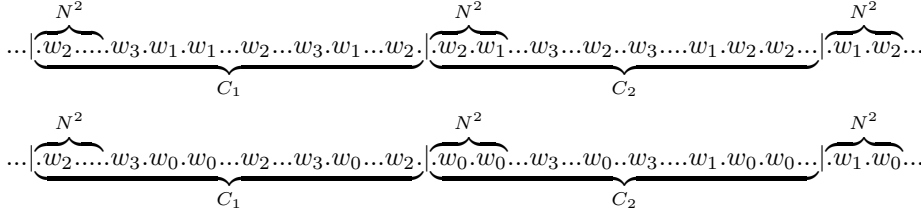


Figure 6: A \mathcal{P} -name before and after modification.

Let C'_i denote the right part of C_i obtained by cutting off its left N^2 entries.

First observe, that the change affects only a subset of $\bigcup_{i=1}^k \bigcup_{m=0}^{l-1} \sigma^{-m}(w_i)$, whose measure is smaller than ϵ . Thus the distance between \mathcal{P} and the partition \mathcal{P}' after the modification is less than ϵ .

Notice also, that the modification completely forbids the word w_i within any C_i and N^2 positions right from it, because all “old” occurrences are removed, and the insertions of the block w_0 do not create any overlapping “new” instances of w_i . On the other hand, these modifications do not affect inside C'_i the words w_j with $j \neq i$ which have not overlapped with w_i before the change.

For fixed $n \in [N, N^2]$ and $i \in [1, k]$ observe an n -block B' (over the partition \mathcal{P}') obtained from a “good” n -block B over \mathcal{P} appearing inside some C'_i . Such blocks (with all possible values of i) still cover more than $1 - \frac{\epsilon}{2} - \frac{2N^2}{p} \geq 1 - \epsilon$ of the

space. Because, for each $j \neq i$, B' contains at least one unaffected copy of w_j (not overlapping with w_i before the change), B' cannot occur with its leftmost position located in any C_j except for $j = i$. On the other hand, inside C'_i it occurs as many times as B did before the change. Because the blocks C'_i jointly contain a fraction $\frac{\frac{p}{2} - N^2}{kp} \geq \frac{1-\epsilon}{2k}$ of all $\frac{p}{2}$ -blocks, only a fraction of at most $\frac{2k\delta}{1-\frac{\epsilon}{2}} = \frac{1}{2}$ of all blocks C'_i may contain less than M copies of B' . Thus the measure $\mu(B')$ of the cylinder B' (with respect to the partition \mathcal{P}') is at least $\frac{M}{2kp} = \frac{t}{p}$. The waiting time for B' is not larger than p only within C_i and the preceding p -block, so $\mu\{V_{B'} \leq p\} \leq \frac{2}{k} < \epsilon$. After normalizing, we obtain $F_{B'}(t) < \epsilon$. We have proved that $\mathcal{P}' \in \mathcal{E}_{\epsilon,t,N}$. This completes the proof of the claim that $\mathcal{E}_{\epsilon,t}$ is dense among the partitions, and ends the whole proof. \square

For a more complete image of a process generated by a typical partition, let us formulate one more fact.

Fact 5. *Let (X, μ, T) be an ergodic measure-preserving transformation of with positive entropy. Fix some $2 \leq m \leq \aleph_0$. Then, in the Polsh space of all measurable partitions \mathcal{P} of X into at most m elements, the set of partitions generating positive entropy is open and dense.*

Proof. It is known that entropy is continuous in the Rokhlin metric, so positive entropy is an open property (see e.g. [P]). To obtain density it suffices to perturb a zero-entropy partition by a small set not measurable with respect to the Pinsker algebra. \square

CONSEQUENCES FOR LIMIT LAWS

The studies of limit laws for return/hitting time statistics are based on the following approach: For $x \in \mathcal{P}^{\mathbb{Z}}$ define $F_{x,n} = F_B$ (and $\tilde{F}_{x,n} = \tilde{F}_B$), where B is the block $x[0, n)$ (or the cylinder in \mathcal{P}^n containing x). Because for nondecreasing functions $F : [0, \infty) \rightarrow [0, 1]$, the weak convergence coincides with the convergence at continuity points, and it makes the space of such functions metric and compact, for every x there exists a well defined collection of limit distributions for $F_{x,n}$ (and for $\tilde{F}_{x,n}$) as $n \rightarrow \infty$. They are called *limit laws for the hitting (return) times at x* . Due to the integral relation ($F_B \approx G_B$) a sequence of return time distributions converges weakly if and only if the corresponding hitting time distributions converge pointwise (see [H-L-V]), so the limit laws for the return times completely determine those for hitting times and *vice versa*. A limit law is *essential* if it appears along some subsequence (n_k) for x 's in a set of positive measure. In particular, the strongest situation occurs when there exists an almost sure limit law along the full sequence (n) . In such case the process is said to have *exponential asymptotics*. Most of the results concerning the limit laws, obtained so far, can be classified in three major groups:

- a) characterizations of possible essential limit laws for specific zero entropy processes (e.g. [D-M], [C-K]; these limit laws are usually atomic for return times or piecewise linear for hitting times),
- b) finding classes of processes with exponential asymptotics (e.g. [A-G], [H-S-V]), and
- c) results concerning non-essential limit laws, limit laws along sets other than cylinders (see [L]; every probabilistic distribution with expected value not exceeding 1 can occur in any process as the limit law for such general return times), or other very specific topics.

As a consequence of our Theorem 1, we obtain, for the first time, a serious bound on the possible essential limit laws for the hitting time statistics along cylinders in the general class of ergodic positive entropy processes. The statement (1) below is even slightly stronger, because we require, for a subsequence, convergence on a positive measure set, but not necessarily to a common limit.

Theorem 3. *Assume ergodicity and positive entropy of the process $(\mathcal{P}^{\mathbb{Z}}, \mu, \sigma)$.*

- (1) *If a subsequence (n_k) is such that \tilde{F}_{x, n_k} converge pointwise to some limit laws \tilde{F}_x on a positive measure set A of points x , then almost surely on A , $\tilde{F}_x(t) \leq 1 - e^{-t}$ at each $t \geq 0$.*
- (2) *If (n_k) grows sufficiently fast, then there is a full measure set, such that for every x in this set holds: $\limsup_k \tilde{F}_{x, n_k}(t) \leq 1 - e^{-t}$ at each $t \geq 0$.*

Proof. The implication from Theorem 1 to Theorem 3 is obvious and we leave it to the reader. For (2) we hint that (n_k) must grow fast enough to ensure summability of the measures of the sets where the intensity of repelling persists, then the Borel-Cantelli Lemma applies. \square

Our Theorem 2 (again combined with the Borel-Cantelli Lemma) shows that a typical positive entropy process (including Bernoulli processes) admits the zero function as an essential limit law for the distributions of the waiting time. In particular, not all Bernoulli processes have exponential asymptotics.

AN EXAMPLE

It is important not to be misled by an oversimplified approach to Theorem 1. The “decay of repelling” in positive entropy processes appears to agree with the intuitive understanding of entropy as chaos: repelling is a “self-organizing” property; it leads to a more uniform, hence less chaotic, distribution of an event along a typical orbit. Thus one might expect that repelling with intensity ϵ revealed by a fraction ξ of all n -blocks contributes to lowering an upper estimate of the entropy by some percentage proportional to ξ and depending increasingly on ϵ . If this happens for infinitely many lengths n with the same parameters ξ and ϵ , the entropy should be driven to zero by a geometric progression. Surprisingly, it is not quite so, and the phenomenon has more subtle grounds. We will present an example which exhibits the incorrectness of such intuition. Note also that in the proof of Theorem 1 the entropy is “killed completely in one step”, that means, positive entropy and persistent repelling lead to a contradiction by examining the blocks of one sufficiently large length n ; we do not use any iterated procedure requiring repelling for infinitely many lengths.

The construction below will show that for each $\delta > 0$ and $n \in \mathbb{N}$ there exists $N \in \mathbb{N}$ and an ergodic process on N symbols with entropy $\log_2 N - \delta$, such that the n -blocks from a collection of joint measure equal to $\frac{1}{n}$ repel with nearly the maximal possible intensity e^{-1} . Because δ can be extremely small compared to $\frac{1}{n}$, this construction illustrates, that there is no “reduction of entropy” by an amount proportional to the fraction of blocks which reveal strong repelling.

Example 1. Let \mathcal{P} be an alphabet of a large cardinality N . Divide \mathcal{P} into two disjoint subsets, one, denoted \mathcal{P}_0 , of cardinality $N_0 = N2^{-\delta}$ and the relatively small (but still very large) rest which we denote by $\{1, 2, \dots, r\}$ (we will refer to these symbols as “markers”). For $i = 1, 2, \dots, r$, let \mathcal{B}_i be the collection of all n -blocks whose first $n - 1$ symbols belong to \mathcal{P}_0 and the terminal symbol is the marker

i. The cardinality of \mathcal{B}_i is N_0^{n-1} . Let \mathcal{C}_i be the collection of all blocks of length nN_0^{n-1} obtained as concatenations of blocks from \mathcal{B}_i using each of them exactly once. The cardinality of \mathcal{C}_i is $(N_0^{n-1})!$. Let X be the subshift whose points are infinite concatenations of blocks from $\bigcup_{i=1}^r \mathcal{C}_i$, in which every block belonging to \mathcal{C}_i is followed by a block from \mathcal{C}_{i+1} ($1 \leq i < r$) and every block belonging to \mathcal{C}_r is followed by a block from \mathcal{C}_1 . Let μ be the shift-invariant measure of maximal entropy on X . It is immediate to see that the entropy of μ is $\frac{1}{nN_0^{n-1}} \log_2((N_0^{n-1})!)$, which, for large N , nearly equals $\log_2 N_0 = \log_2 N - \delta$. Finally observe that the measure of each $B \in \mathcal{B}_i$ equals $\frac{1}{nrN_0^{n-1}}$, the joint measure of $\bigcup_{i=1}^r \mathcal{B}_i$ is exactly $\frac{1}{n}$, and every block B from this family appears in any $x \in X$ with gaps ranging between $\frac{1-\frac{1}{n}}{\mu(B)}$ and $\frac{1+\frac{1}{n}}{\mu(B)}$, revealing strong repelling.

Remark 1. Viewing the blocks of length nrN_0^{n-1} starting with a block from \mathcal{C}_1 as a new alphabet, and repeating the above construction inductively, we can produce an example (with the measure of maximal entropy on the intersection of systems created in consecutive steps) with entropy $\log_2 N - 2\delta$, in which the strong repelling will occur with probability $\frac{1}{n_k}$ for infinitely many lengths n_k .

Remark 2. The process described in the above remark is (somewhat coincidentally; it was not designed for that) bilaterally deterministic: for every $m \in \mathbb{N}$ the sigma-field $\mathcal{P}^{(-\infty, -m] \cup [m, \infty)}$ equals the full (product) sigma-field. Indeed, suppose we see all entries of a \mathcal{P} -name of a point x except on the interval $(-m, m)$. In a typical point, this interval is contained between a pair of successive markers i for some level k of the inductive construction. Then, by examining this name's entries far enough to the left and right we will see complete all but one (the one covering the coordinate zero) blocks from the family \mathcal{B}_i which constitute the block $C \in \mathcal{C}_i$ covering the considered interval. Because every block from \mathcal{B}_i is used in C exactly once, by elimination, we will be able to determine the missing block from \mathcal{B}_i and hence all symbols in $(-m, m)$.

QUESTIONS

Question 1. Is there a speed of the convergence to zero of the joint measure of the “bad” blocks in Theorem 1? More precisely, does there exist a positive function $s(n, \epsilon, \#\mathcal{P})$ converging to zero as n grows, such that if for some ϵ and infinitely many n 's, the joint measure of the n -blocks which repel with intensity ϵ exceeds $s(n, \epsilon, \#\mathcal{P})$, then the process has necessarily entropy zero? (By the Example 1, $\frac{1}{n}$ is not enough.)

Question 2. Can one strengthen the Theorem 3 as follows:

$$\limsup_{n \rightarrow \infty} \tilde{F}_{x,n} \leq 1 - e^{-t} \quad \mu\text{-almost everywhere?}$$

Question 3. In Lemma 3, can one obtain \mathcal{P}^r conditionally β -independent of jointly the past and all return times $R_B^{(k)}$ ($k \geq 1$) (for sufficiently large n , with μ -tolerance β for $B \in \mathcal{P}^{-n}$)? In other words, can the β -independent process $((\mathcal{P}_B^r)^\mathbb{Z}, \mu_B, \sigma_B)$ be obtained β -independent of the factor-process generated by the partition into B and its complement?

Question 4. (suggested by J-P. Thouvenot) Find a purely combinatorial proof of Theorem 1, by counting the quantity of very long strings (of length m) inside

which a positive fraction (in measure) of all n -blocks repel with a fixed intensity. For sufficiently large n this quantity should be eventually (as $m \rightarrow \infty$) smaller than h^m for any preassigned positive h .

REFERENCES

- [A-G] Abadi, M. and Galves, A., *Inequalities for the occurrence times of rare events in mixing processes. The state of the art. Inhomogeneous random systems*, Markov Process. Related Fields **7** (2001), 97–112.
- [C-K] Chaumoitre, V. and Kupsa, M., *Asymptotics for return times of rank one systems*, Stochastics and Dynamics **5** (2005), 65–73.
- [C] Coelho, Z., *Asymptotic laws for symbolic dynamical systems*, Topics in symbolic dynamics and applications, London Math. Soc. Lecture Note Ser., vol. 279, Cambridge Univ. Press, 2000, pp. 123–165.
- [D-M] Durand, F. and Maass, A., *Limit laws for entrance times for low-complexity Cantor minimal systems*, Nonlinearity **14** (2001), 683–700.
- [G] Gurevič, B. M., *One- and two-sided regularity of stationary random processes*, Dokl. Akad. Nauk SSSR **210** (1973), 763–766.
- [H-L-V] Haydn, N., Lacroix, Y. and Vaienti, S., *Hitting and return times in ergodic dynamical systems*, Ann. Probab. **33** (2005), 2043–2050.
- [H-S-V] Hirata, N. Saussol, B. and Vaienti, S., *Statistics of return times: A general framework and new applications*, Commun. Math. Phys. **206** (1999), 33–55.
- [Kc] Kac, M., *On the notion of recurrence in discrete stochastic processes*, Bull. Amer. Math. Soc. **53** (1947), 1002–1010.
- [Km] Kammerer P., *Das Gesetz der Serie, eine Lehre von den Wiederholungen im Lebens und im Weltgeschehen*, Stuttgart und Berlin, 1919.
- [L] Lacroix, Y., *Possible limit laws for entrance times of an ergodic aperiodic dynamical system*, Israel J. Math. **132** (2002), 253–263.
- [Mi] Von Mises, *Probability, Statistics and Truth*, 2nd ed. rev., New York, Dover, 1981.
- [O-W1] Ornstein, D.S. and Weiss, B., *Every transformation is bilaterally deterministic*, Israel J. Math. **21** (1975), 154–158.
- [O-W2] Ornstein, D.S. and Weiss, B., *Entropy and recurrence rates for stationary random fields*, IEEE Trans. Inform. Theory **48** (2002), 1694–1697.
- [P] Petersen, K., *Ergodic Theory*, Cambridge Univ. Press, Cambridge, 1983.
- [Ro] Rokhlin, V.A., *Selected topics from the metric theory of dynamical systems*, Am. Math. Soc., Transl. **2** (1966), 171–240.
- [Ru] Rudolph, D.J., *If a two-point extension of a Bernoulli shift has an ergodic square, then it is Bernoulli*, Israel J. Math. **30** (1978), 159–180.
- [Sh] Shields, P., *The theory of Bernoulli shifts*, Chicago Lectures in Mathematics, The University of Chicago Press, Chicago, 1973.
- [Sm] Smorodinsky, M., *On Ornstein's isomorphism theorem for Bernoulli shifts*, Advances in Math. **9** (1972), 1–9.
- [Wa] Walters, P., *Ergodic theory—introductory lectures*, Lecture Notes in Mathematics, vol. 458, Springer-Verlag, Berlin, 1975.
- [Wi] *Wikipedia*, http://en.wikipedia.org/wiki/Charles_Wells_%28gambler%29.

INSTITUTE OF MATHEMATICS AND COMPUTER SCIENCE, WROCLAW UNIVERSITY OF TECHNOLOGY, WYBRZEŻE WYSPIAŃSKIEGO 27, 50-370 WROCLAW, POLAND
E-mail address: downar@pwr.wroc.pl

INSTITUT DES SCIENCES DE L'INGÉNIEUR DE TOULON ET DU VAR, AVENUE G. POMPIDOU, B.P. 56, 83162 LA VALETTE DU VAR CEDEX, FRANCE
E-mail address: yves.lacroix@univ-tln.fr